FDA Discussion Questions
CADe Radiological Panel Meeting
November 18, 2009

1. Considering the input provided in the March 2008 Panel meeting, has the Agency adequately addressed in these two draft documents the major points of discussion and recommendations for premarket review (comparison of the device description, device standalone performance testing, clinical performance testing, and labeling)? Describe any areas of concern that should be clarified. Identify and describe areas that should be modified, removed, or added, and provide your rationale for those changes.

2. Under Section 6 of the 510(k) draft guidance, the Agency states that a clinical performance assessment will usually be necessary to demonstrate substantial equivalence to a predicate CADe device. A clinical performance assessment is expected for all original PMAs. The clinical performance guidance was developed to provide recommendations for designing a reader study to support either a 510(k) or PMA.

    a. Please discuss what you consider to be a valid control arm for such studies.

        i. What should be the expected clinically meaningful outcome to demonstrate that a new or modified CADe device is substantially equivalent to a legally marketed predicate CADe device?

        ii. What should be the expected outcome to demonstrate a reasonable assurance of safety and effectiveness for a CADe device subject to a PMA or PMA supplement?

As a means for discussion we have provided the following examples:

        iii. A manufacturer has a legally marketed CADe device and wants to submit a new 510(k) for an upgraded version. They are planning their clinical performance assessment. Should the control arm of their study be image reading without CADe or image reading on their previously cleared CADe device (the predicate for the new 510(k))? What should the statistical hypothesis be? If the control arm is image reading without CADe, what level of superiority is needed against an unaided read to demonstrate substantial equivalence to other legally marketed CADe devices for the same intended use but with different technological characteristics (e.g., different algorithms)? If the control arm is against their previously cleared CADe device, what is a clinically acceptable delta to demonstrate non-inferiority?

iv. A manufacturer has a legally marketed CADe device, which falls into Class III, with an approved PMA. They are planning to submit a PMA supplement to support an update to their device. Should the control arm for their clinical performance assessment be their original device or image reading without CADe? Do the same conditions for the statistical hypothesis hold true as you described for (iii)?

b. Please describe under what conditions the Agency should consider accepting standalone performance in lieu of clinical performance data for a CADe device? As a means for discussion we have provided the following examples for consideration:

i. A manufacturer has a new CADe device and has done standalone testing comparing it to an already cleared CADe device. The standalone performance for both the new device and the predicate device were derived from the same database of cases and using the same truthing and scoring methodologies. The new CADe identifies additional abnormalities that are not detected by the predicate device, but misses some of the abnormalities that were detected by the predicate device. Additionally, the new CADe had fewer false positive marks than the predicate device. In this situation, would standalone performance testing be sufficient to demonstrate substantial equivalence or should the manufacturer perform a clinical performance assessment of their new CADe device? Please describe your rationale for your answer.

ii. Similarly to (i), except that the standalone performance of the new device was derived with either a different database of cases or with different truthing and scoring methodologies, providing for a less direct comparison between the two devices. In this situation, should the manufacturer perform a clinical performance assessment of their new CADe device?

iii. A manufacturer's new CADe runs the algorithm from their old CADe but applies to the output an additional algorithm designed to mask specific types of false positives. In standalone testing, the masking did not eliminate any of the cancers. Should the manufacturer perform a clinical performance assessment of their new CADe device?

iv. A manufacturer previously received clearance for a CADe device that can serve as a predicate device. They intend to provide to the intendeduser a different prompt format from that of their cleared CADe device (e.g., findings are now marked with a circle rather than an arrow). The prompt format is the only change made; the

CADe algorithms were not changed or modified in any manner. Should they perform another clinical performance assessment?

v. A manufacturer previously received clearance for a CADe device that can serve as a predicate device. They intend to seek FDA clearance of an updated device (i.e., update of the algorithms) since they improved device standalone performance (e.g., higher sensitivity and less false positives detection) on both the training database and validation database. Does an improvement in device performance translate to improvement in clinician performance? Should a clinical performance assessment be performed to show that the intended user performance is non-inferior with the new device as compared to the predicate device?

3. Under Section 4 for the Clinical Performance Assessment draft guidance, the Agency describes considerations regarding the use of sample enrichment, study endpoints, reader characteristics. Has the Agency provided sufficient clarity of its expectations for what constitutes a scientifically sound study? To assist in the discussion we have provided the following questions and examples.

a. In a 510(k) submission to support clearance of a CADe device, are there circumstances where test data can be reused in (1) standalone assessment or (2) clinical assessment? If so, what types of constraints do you recommend on this reuse of data?

For example, in a CADe 510(k) device submission, the manufacturer sequestered the test data set and only used it once to support clearance of a CADe device. This CADe device is now the predicate in a 510(k) submission of the same manufacturer's new device. Can test data be reused to support clearance of the new CADe device? If so what are the constraints you recommend on this reuse?

b. The guidance calls for the trial readers to be "representative of the intended population of clinical users." Can you provide examples of sets of readers that are representative of clinical users? Should there be a minimum number of readers? If there are important subgroups of readers, should the number of readers in each of the subgroups be proportional to the numbers in the population of clinical users?

c. A manufacturer's CADe device is designed to detect abnormalities on mammograms. Is there a minimum number of cancers that should be included in their clinical study to ensure that the entire spectrum of cancer is represented? Should their clinical performance assessment be powered so that statistically significant results can be obtained for the clinically relevant subgroups of cancer manifesting as microcalcification clusters and cancers manifesting as masses? Does the answer depend on whether

or not we have prior experience that CAD devices do not perform well in one of the subgroups, e.g., masses?

d. Would your answers to item (c) apply to other CADe devices? For example, a CADe device is designed to detect lung nodules. Should their clinical performance assessment be powered so that statistically significant results can be obtained for the clinically relevant subgroups of lesions, for example, nodules near the mediastinum vs. the peripheral lung fields? Or would this only be expected if the manufacturer proposed to make such claims in their labeling?

e. Manufacturers typically report Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), Sensitivity (Se) and Specificity (Sp) for their clinical performance studies. Should the studies be powered for all summary endpoints? If not, which endpoint(s) should be used to size (power) the study? For example, a clinical performance assessment for a breast CADe device could be powered for AUC based on a radiologist's reported probability that an image contains a malignancy, or it could be powered for sensitivity and specificity based on a cut point (e.g., 3 or 4) in the BIRADS scale.

f. A manufacturer has developed a breast CADe device and would like to make the claim that their CADe device will help detect breast cancer earlier than an unaided read. What would be the appropriate endpoint to use in their clinical performance study to support this claim, recognizing that the prevalence of breast cancer is low?

4. These two draft guidance documents, when finalized, will represent a change from our past approach and thought process concerning the performance data requirements for CADe devices. Many CADe devices are currently on the market, as are a wide range of medical device equipment for generating images to which CADe is applied. Please discuss the conditions in which clinical performance assessments should be conducted for devices under review for the first time, i.e., for devices new or previously cleared with changes, to provide adequate assurance that the CADe performance data are generalizable across medical imaging devices. For the purpose of discussion we have provided the following questions and examples.

a. A manufacturer's CT CADe device is intended to be used on a variety of CT devices. There are a large number of CT systems currently on the market. How should a clinical study be designed to demonstrate that the CADe performance is generalizable across all CT systems? Should the study design include every type of CT system with which the CADe device is intended to be used?

b. A manufacturer's colon CADe device can be used for both 2D and 3D

interpretation. How should a clinical study be designed to assess CADe performance given the variability of how physicians may use the 2D and 3D modes?

c. A manufacturer has a new breast CADe device and would like to market it for use with all legally marketed Full Field Digital Mammography (FFDM) systems. How should the clinical study be designed to demonstrate that the CADe performance is generalizable across all legally marketed FFDM? Should clinical studies with each legally marketed FFDM be required?

d. A manufacturer has a breast CADe device approved for use with a specific legally marketed Full Field Digital Mammography (FFDM) based on a robust MRMC study. They would like to market it for use with an additional legally marketed FFDM. Is a clinical performance assessment (i.e., reader study) necessary to assess the CADe for use with the new FFDM or is standalone performance data sufficient to demonstrate comparable performance based on the specifications of the device?

5. The following questions seek additional discussion and clarification of specific responses received at the March 2008 panel meeting.

a. Manufacturers often make modifications to their devices that could be considered “minor”. The Agency is seeking your input on the significance of “minor modifications” to CADe devices. Is there a clear definition of what would constitute a “minor modification”? Can you identify what “minor modifications”, if any, would not need reader studies to establish a reasonable assurance of safety and effectiveness for a PMA submission? Can you identify what “minor modifications”, if any, would not need reader studies to establish substantial equivalence for a 510(k) submission?

b. Mammographic CADe devices contain separate and distinct algorithms that detect masses versus microcalcifications. The following questions seek input on whether this distinction should have regulatory significance:

i. If a regulatory submission for an original mammography CADe device reveals that reader performance does not show safety and effectiveness separately for masses and microcalcifications, e.g., suppose safety and effectiveness is shown for microcalcifications but not for masses, should the indications for use specify that the device is only indicated for the detection of microcalcifications? If yes, do you believe that the mass detection portion of the device should be disabled/removed?

ii. For devices already on the market, do you believe the Agency

should address this issue by asking manufacturers to provide data demonstrating safety and effectiveness for mass detection, and if manufacturers are unable to do so, should they modify the labeling of their devices to reflect this fact?

c. Mammography CADe devices are currently labeled as "second readers." Do you believe that these devices are used in a "second-read" mode by the majority of radiologists who use the devices in clinical practice? If not, is this an important issue the Agency should address through a regulatory means?

d. Do you believe CADe labeling should address reading time and if so how?

e. Do you believe the draft guidance documents adequately explain the clinical meaning of the area under the ROC curve? Do you believe that the draft guidance documents adequately reflect the use of alternative performance metrics?

6. Historically, PMA applications for mammography CADe devices included retrospective studies with enriched data but did not include data from prospective clinical trials due to the significant burden of adequately powering a prospective study. Published literature of clinical studies evaluating CAD Mammography in the postmarket setting have not presented a consensus on findings or have limitations that minimize generalizability. Please comment on the following:

   a. Although a retrospective study with enriched data may be adequate to demonstrate a reasonable assurance of safety and effectiveness, should the question of device performance under actual conditions of use (postmarket) be answered by a post-approval study?

7. The Agency seeks the input and advice from the panel regarding the interpretation of clinical studies of CADe use. Data from published meta-analyses point to an increase in the recall rate and biopsy rate emanating from mammography CADe usage (57, 79)

   a. Please discuss what threshold is appropriate in clinical settings to balance increased cancer detection against recall rates leading to biopsy and additional surgery. Are there additional data elements to consider?

8. Currently, FDA has approved mammography CADe devices and one chest x-ray CADe device as Class III devices; lung CT and colon CT CADe devices have been cleared as Class II devices. As explained above in Section II, classification of devices generally reflects risk. Are there risks unique to mammography and chest x-ray CADe devices that justify their continued regulation in a separate class from other CADe devices? Alternatively, are the risks for CADe devices for

lung CT and colon CT similar to mammography and chest x-ray CADe to warrant their reclassification to Class III? Please describe the risks associated with these different CADe applications with respect to either their unique risks or similar risks with respect to the Agency's reclassification consideration.

9. FDA has the authority to require postmarket studies for some devices (though this authority is, generally speaking, more readily available for Class III than for Class II devices). Are there long-term questions about the performance of CADe devices that will remain unanswered by premarket clinical assessments and that should be answered by postmarket studies? If so, in what instances should a postmarket study be used to address unresolved questions or findings generated by the premarket clinical study? Are there any particular sub-groups or other considerations for which this is especially important and should be considered?